(12) PATENT ABRIDGMENT (11) Document No. AU-B-46468/85
(19) AUSTRALIAN PATENT OFFICE (10) Acceptance No. 592062



(54) Title
**SYNTHETIC REGULATION REGION**

International Patent Classification(s)
(51)[4] **C12N 015/00 C07G 017/00 C12N 001/20 C12P 019/34 C12P 021/00**

(21) Application No. : **46468/85** (22) Application Date : **20.08.85**

(30) Priority Data

| (31) Number | (32) Date | (33) Country |
|---|---|---|
| 3430683 | 21.08.84 | DE FEDERAL REPUBLIC OF GERMANY |

(43) Publication Date : **27.02.86**

(44) Publication Date of Accepted Application : **04.01.90**

(71) Applicant(s)
**HOECHST AKTIENGESELLSCHAFT**

(72) Inventor(s)
**JOACHIM ENGELS; EUGEN UHLMANN; MICHAEL LEINEWEBER**

(74) Attorney or Agent
**WATERMARK MELBOURNE**

(56) Prior Art Documents
**AU 83760/82 553901 C12N 15/00 C12N 1/20 C12R 1/19**
**AU 32355/84 C12N 15/00**
**AU 23239/84 C12N 15/00**

(57) Claim

1. A synthetic regulation region for the expression of heterologous genes in E. coli, which contains a promoter, a modified lac operator and a ribosomal binding site, having one or more of the following features:

a) a spacer group of 15 to 18 base-pairs is located between the -35 and the -10 regions, and

b) a spacer group of 6 to 14 base-pairs is located between the ribosomal binding site and the ATG start codon

and optionally:

c) the -35 region in the promoter has the following nucleotide sequence (coding strand)

TTGACAT or CTTGACAT,

d) the -10 region in the promoter has the following nucleotide sequence (coding strand)

GTATA..

COMMONWEALTH OF AUSTRALIA 592062

*Patents Act 1952-1969*

# CONVENTION APPLICATION FOR A PATENT

LODGED AT SUB-OFFICE
20 AUG 1985
Melbourne

(1) Here insert (in full) Name or Names of Applicant or Applicants, followed by Address (es).

~~We~~ / We (1) HOECHST AKTIENGESELLSCHAFT,
of 45 Bruningstrasse, D-6230 Frankfurt/Main 80,
Federal Republic of Germany

(2) Here insert Title of Invention.

hereby apply for the grant of a Patent for an invention entitled: (2)
SYNTHETIC REGULATION REGION

(3) Here insert number(s) of basic application(s)

which is described in the accompanying complete specification. This application is a Convention application and is based on the application numbered (3)
P34 30 683.8

(4) Here insert Name of basic Country or Countries, and basic date or dates

for a patent or similar protection made in (4) Federal Republic of Germany on 21st August 1984

APPLICATION ACCEPTED AND AMENDMENTS ALLOWED 24-10-85

~~My~~ / Our address for service is Messrs. Edwd. Waters & Sons. Patent Attorneys. 50 Queen Street, Melbourne, Victoria, Australia.

DATED this 19th day of August 19 85.

(5) Signature (s) of Applicant (s) or Seal of Company and Signatures of its Officers as prescribed by its Articles of Association.

(5) HOECHST AKTIENGESELLSCHAFT

by James Murray

James Murray
Attorney

To:
THE COMMISSIONER OF PATENTS.

Form 7

COMMONWEALTH OF AUSTRALIA

Patents Act 1952

DECLARATION IN SUPPORT OF A CONVENTION APPLICATION UNDER PART XVI. FOR A PATENT.

---

In support of the Convention application made under Part XVI. of the Patents Act 1952 by HOECHST AKTIENGESELLSCHAFT of 45, Brüningstrasse, D-6230 Frankfurt/Main 80, Federal Republic of Germany for a patent for an invention entitled:

"SYNTHETIC REGULATION REGION"

We, Johann-Heinrich Reuter, 4 Bodenheimer Straße, D-6500 Mainz, Franz Lapice, 2 Sandweg, D-6233 Kelkheim (Taunus); Federal Republic of Germany

do solemnly and sincerely declare as follows:

1. We are authorized by HOECHST AKTIENGESELLSCHAFT the applicant for the patent to make this declaration on its behalf.

2. The basic application as defined by Section 141 of the Act was made in the Federal Republic of Germany under No. P 34 30 683.8 on August 21, 1984 by HOECHST AKTIENGESELLSCHAFT

3. a) Joachim Engels, 1 Feldbergstraße, D-6242 Kronberg/Taunus
b) Eugen Uhlmann, 287 Washington Street, Belmont, Mass. 02178
c) Michael Leineweber, 74 Heimchenweg, D-6230 Frankfurt am Main 80
a) and c) Federal Republic of Germany, b) United States of America

is/are the actual inventor(s) of the invention and the facts upon which HOECHST AKTIENGESELLSCHAFT is entitled to make the application are as follows:

The said HOECHST AKTIENGESELLSCHAFT is the assignee of the said Joachim Engels, Eugen Uhlmann and Michael Leineweber

4. The basic application referred to in paragraph 2 of this Declaration was the first application made in a Convention country in respect of the invention the subject of the application.

DECLARED at Frankfurt/Main, Federal Republic of Germany

this 11th day of July 1985

To the Commissioner of Patents

Hoechst Aktiengesellschaft

Prokurist (ppa.Reuter) Authorized signatory (i.V.Lapice)

PAT 510

592062

Form 10

COMMONWEALTH OF AUSTRALIA

PATENTS ACT 1952-69

# COMPLETE SPECIFICATION

(ORIGINAL)

Class Int. Class

Application Number:
Lodged:

Complete Specification Lodged:
Accepted:
Published:

Priority :

Related Art :

This document contains the amendments made under Section 49 and is correct for printing.

46468/85

Name of Applicant : HOECHST AKTIENGESELLSCHAFT

Address of Applicant : 45 Bruningstrasse, D-6230 Frankfurt/Main 80, Federal Republic of Germany

Actual Inventor: JOACHIM ENGELS, EUGEN UHLMANN and MICHAEL LEINEWEBER

Address for Service : EDWD. WATERS & SONS, 50 QUEEN STREET, MELBOURNE, AUSTRALIA, 3000.

Complete Specification for the invention entitled:

SYNTHETIC REGULATION REGION

The following statement is a full description of this invention, including the best method of performing it known to :- US

In the preparation of eukaryotic polypeptides in bacteria, in particular E. coli, by genetic engineering, the "heterologous" gene which codes for the desired eukaryotic polypeptide is incorporated in a suitable vector, and this hybrid vector is then introduced into the bacterial host. However, a number of conditions must be met for this heterologous gene to be able to bring about the production of the desired polypeptide. An essential condition is a functioning regulation region, consisting of an operator, a promoter and a so-called "Shine-Dalgarno sequence", also called an SD sequence or, simplifying (since only the corresponding sequence of the mRNA binds to the ribosome), called the "ribosomal binding site" below.

Correct expression of the gene, that is to say the production of the desired polypeptide, is conditional on recognition of the promoter by the bacterial host. The enzyme RNA polymerase which is intrinsic to the host recognizes a part sequence in the DNA of the promoter and binds to this part sequence. This brings about an opening of the double-stranded DNA in this region, whereupon the synthesis of the mRNA on the coding strand (transcription strand) starts.

The operator, which frequently overlaps with the promoter, is recognized by a repressor protein which is intrinsic to the host. The frequency of transcription is controlled by the more or less effective binding of this repressor protein to the operator. This system is affected by inducers (inducer molecules) which bind the repressor protein and thus activate the operator.

Finally, the ribosomal binding site is responsible for the initially synthesized part of the mRNA containing an RNA sequence for the binding to the ribosome on which the translation into the polypeptide takes place.

Thus the regulation region is responsible for the expression of the gene to give the desired polypeptide via

the transcription stage (transcription of DNA into mRNA) and the subsequent translation. Apart from the sequence of nucleotides, an important point about the structure of a regulation region of this type is the geometry, that is to say the spatial arrangement of the promoter, operator and ribosomal binding site.

In the following text, the numbering of the nucleotides relates to the site at which transcription starts (zero), counting being, as usual, from 5' to 3'.

Natural promoters for E. coli RNA polymerase have two regions of DNA sequences which are preserved. One is the -35 region and the other is the -10 region, also called the "Pribnow-Schaller box", the numbering being based on the abovementioned nucleotide numbering, that is to say these regions are located upstream of the start of transcription.

The regulation sequence according to the invention represents a modification of the natural regulation sequences and ensures optimal binding of the RNA polymerase to the promoter and effective utilization of the operator. The regulation sequence according to the invention can either be placed directly upstream of the heterologous gene, whereupon the desired polypeptide (with methionine at the amino terminal end) is expressed, or a bacterial gene - in whole or in part - is interpolated upstream of the heterologous gene, this leading to expression of a fusion protein in which (a portion of) the bacterial protein is bonded to the amino terminal end of the desired polypeptide.

The synthetic regulation region according to the invention for the expression of heterologous genes in E. coli, containing a promoter, a modified lac operator and a ribosomal binding site, comprises one or more of the following features:

~~a) the -35 region in the promoter has the following nucleotide sequence (coding strand)~~

~~TTGACA,~~

a) a spacer group of 15 to 18 base-pairs is located between the -35 and the -10 regions, and

b) a spacer group of 6 to 14 base-pairs is located between the ribosomal binding site and the ATG start codon

and optionally:

c) the -35 region in the promoter has the following nucleotide sequence (coding strand)

TTGACAT or CTTGACAT,

d) the -10 region in the promoter has the following nucleotide sequence (coding strand)

GTATAAT.

Other embodiments of this invention include the following:

A regulation region wherein the spacer region has 15 base-pairs (bp) and/or is A-T rich. A regulation region wherein the spacer group is between the ribosome binding site (r.b.s.) and the start codon has 10 bp and/or is A-T rich. A regulation region wherein the r.b.s. is rich in purines. A regulation region wherein the modified lac operon has the DNA sequence I or IIa as herein defined.

Apart from the advantages already mentioned, the regulation region according to the invention is distinguished by being very variable and, because there is a number of unique restriction sites, it allows the individual elements, namely the promoter, operator and ribosomal binding site, to be cut out and combined with known systems. Moreover, by modification of the spacer groups, it is possible to vary the geometry and suit it even better to the circumstances of the individual case.

The regulation region according to the invention is preferably constructed by complete synthesis. It is possible to use the known DNA synthetic methods for this purpose, for example the phosphite method.

DNA sequence I (Appendix) shows an advantageous embodiment of the total regulation region. DNA sequences IIa to IIh show specific, preferred embodiments of sequence I. For the synthesis of sequences I and II, a few nucleotide pairs which permit attack by restriction enzymes for "cutting to size" are additionally attached to the 5' and 3' ends in each case. DNA sequence IIa is also given in complete detail, three nucleotide pairs being placed on each of the 5' and 3' ends. Obviously, it would also be possible to attach other or more nucleotide pairs to these.

In the Examples which follow, specific embodiments of the invention are illustrated in detail, from which the large number of possible modifications and combinations is evident to those skilled in the art. Unless otherwise specified, percentages data in these Examples relate to weight.

Example 1

Synthesis of DNA sequence IIa

a) Chemical synthesis of a single-stranded oligonucleotide

The synthesis of the structural units of the gene is illustrated by the example of structural unit Ia of the gene, which comprises nucleotides 1-19 (plus three others at the 5' end to allow attack by Bam HI) of the coding strand. Using known methods (M. J. Gait et al., Nucleic Acids Res. 8 (1980) 1081-1096), the nucleoside located at the 3' end, that is to say in the present case thymidine (nucleotide no. 19), is covalently bonded via the 3'-hydroxyl group to silica gel ($^{(R)}$FRACTOSIL, supplied by Merck). For this purpose, first the silica gel is reacted with 3-(triethoxysilyl)propylamine with elimination of ethanol and formation of an Si-O-Si bond. The thymidine is reacted with the modified carrier in the presence of paranitrophenol and N,N'-dicyclohexylcarbodiimide, the free carboxyl group of the succinoyl group acylating the amino radical of the propylamino group.

In the following synthetic steps, the base component is used in the form of the dialkylamide or chloride of the monomethyl ester of the 5'-O-dimethoxytritylnucleoside 3'-phosphorous acid, the adenine being in the form of the $N^6$-benzoyl compound, the cytosine being in the form of the $N^4$-benzoyl compound, the guanine being in the form of the $N^2$-isobutyryl compound and the thymidine, which contains no amino group, being without a protective group.

100 mg of the polymeric carrier which contains 4 µmol of bound thymidine are treated consecutively with the following agents:

a) nitromethane
b) saturated zinc bromide solution in nitromethane containing 1% water
c) methanol
d) tetrahydrofuran
e) acetonitrile
f) 80 µmol of the appropriate nucleoside phosphite and 400 µmol of tetrazole in 1 ml of anhydrous acetonitrile (5 minutes)

g) 20% acetic anhydride in tetrahydrofuran containing 40% lutidine and 10% dimethylaminopyridine (2 minutes)

h) tetrahydrofuran

i) tetrahydrofuran containing 20% water and 40% lutidine

j) 3% iodine in collidine/water/tetrahydrofuran in the ratio by volume of 5:4:1

k) tetrahydrofuran and

l) methanol.

In this context, the term "phosphite" is to be understood to be the monomethyl ester of the deoxyribose 3'-monophosphorous acid, the third valency being saturated by chlorine or a tertiary amino group, for example a morpholino radical. The yields in the individual synthetic steps can in each case be determined, after the detritylation reaction (b), by spectrophotometry by measuring the absorption of the dimethoxytrityl cation at a wavelength of 496 nm.

After the synthesis of the oligonucleotide is complete, the methyl phosphate protective groups of the oligomer are eliminated using p-thiocresol and triethylamine. The oligonucleotide is then removed from the carrier by treatment with ammonia for 3 hours. Treatment of the oligomers with concentrated ammonia for 2 to 3 days quantitatively eliminates the amino protective groups on the bases. The crude product thus obtained is purified by high-pressure liquid chromatography (HPLC) or by polyacrylamide gel electrophoresis.

The other structural units Ib-Ih of the gene, whose nucleotide sequences are derived from DNA sequence IIa, are also synthesized entirely correspondingly.

b) Enzymatic linkage of the single-stranded oligonucleotides

For the phosphorylation of the oligonucleotides at the 5' terminal end, 1.0 nmol of each of oligonucleotides Ib and Ic are treated, with the addition of 10 nmol of adenosine triphosphate, with 6 units of T4 polynucleotide kinase in 20 µl of 50 mM tris HCl buffer (pH 7.6), 10 mM magnesium chloride and 10 mM dithiothreitol (DTT) at 37°C for 30 minutes (C.C. Richardson, Progress in

Nucl. Acid. Res. 2 (1972) 825). The enzyme is inactivated by heating at 95°C for 5 minutes.

Oligonucleotides If and Ig are phosphorylated analogously.

Oligonucleotides Ia and Id and phosphorylated oligonucleotides Ib and Ic are heated in 40 µl of 50 mM tris HCl buffer at 95°C for 5 minutes, and they are then allowed to cool slowly to room temperature. To this mixture are added 20 mM magnesium chloride, 10 mM DTT and 1 mM ATP, and reaction with 100 units of T4 DNA ligase is allowed to continue at 25°C for 16 hours.

Fragments Ie and Ih are linked with phosphorylated fragments If and Ig analogously.

The product of the ligase reaction of oligonucleotides Ia to Id (gene fragment A) is freeze-dried and incubated in 100 µl of a buffer solution (150 mM NaCl, 10 mM tris HCl, pH 7.6, 6 mM magnesium chloride), which contains 200 units of the endonuclease Bam HI, at 37°C for 3 hours.

The product of the ligase reaction of oligonucleotides Ie to Ih (gene fragment B) is freeze-dried and incubated in 100 µl of a buffer solution (100 mM tris HCl, pH 7.5, 50 mM NaCl, 5 mM magnesium chloride), which contains 200 units of endonuclease Eco RI, at 37°C for 3 hours. After the enzyme digestion has been stopped by heating at 95°C for 2 minutes, the cut gene fragments A and B are purified by gel electrophoresis on 15% polyacrylamide gel (without addition of urea, 20 x 40 cm, 2 mm thick), the marker substance used being pBR 322 (supplied by Biolabs) cut with Hae III. After extraction of the DNA bands and purification on (R)Sephadex G50 and Sep Pak (supplied by Waters), gene fragments A and B are linked by "blunt end ligation". A synthetic regulation region corresponding to DNA sequence IIa which has, on the 5' end of the coding strand, an extension for attack by Bam HI and, on the 3' end of the coding strand, an extension for attack by Eco RI is obtained in this way.

Example 2:

Hybrid plasmids which contain the synthetic control region

a) Incorporation of the control region in pUC 8

The commercially available plasmid pUC 8 is opened in known manner using restriction endonucleases Eco RI and Bam HI, and is separated, using 1% low-melting agarose gels, from the oligonucleotides which have been cut out. The cut plasmid is recovered after dissolving the gel at elevated temperature in accordance with the statements of the manufacturers. 1 µg of the pUC 8 plasmid thus opened is ligated with 10 µg of the synthetic control region using T4 DNA ligase at 14°C overnight. In this way, a modified pUC 8 plasmid having the integrated control region is obtained. This hybrid plasmid is represented in Figure 1, in which the control region is indicated by SIP, which stands for "synthetic idealized promoter".

b) Transformation

The strain E. coli K 12 is made competent by treatment with a 70 mM calcium chloride solution, and a suspension of the hybrid plasmid in 10 mM tris HCl buffer (pH 7.5), which is 70 mM in calcium chloride, is added. The transformed strains are selected for ampicillin resistance, and the inserted sequence is verified by Maxam-Gilbert sequence analysis (A. Maxam and W. Gilbert, Proc. Natl. Acad. Sci. USA 74, 560-564 (1977)).

Example 3:

Expression plasmids which contain the synthetic control region

The commercially available plasmid pBR 322 is opened using the restriction enzymes Bam HI and Sal I and is purified on a 1% agarose gel as described above. The synthetic control region is reisolated from the modified pUC 8 derivative by cutting with the enzymes Eco RI and Bam HI, purified on 10% polyacrylamide gels, and recovered by subsequent electroelution. In an analogous manner, the γ-interferon gene is isolated from appropriate hybrid plasmids by cutting with the restriction enzymes Eco RI and Sal I and purifying on a 2% low-melting agarose gel. The hybrid plasmid containing the γ-interferon gene which is used is the plasmid pMX 2 whose preparation is described in German Patent Application

P 34 09 966.2. However, the plasmid described in European Patent Application 0,095,350 can also be used.

The linearized plasmid pBR 322, the synthetic control regions and the γ-interferon gene are then ligated in known manner, the plasmid as shown in Figure 2 being obtained.

Example 4:

Comparison of the activity of the synthetic control region with that of the known potent tac control region.

The plasmid pKK 177.3 known from the literature (E. Amann et al., Gene 25, 167 (1983)) is cut using the restriction enzymes Eco RI and Sal I as described above, and the γ-interferon gene is incorporated, by which means the latter is coupled to the tac control region. The plasmid shown in Figure 3 is obtained.

To remove the tac control region from pKK 177.3, the latter is digested with Bam HI and Sal I. The plasmid thus linearized is ligated with the γ-interferon gene and the synthetic control region as described above. The hybrid plasmid shown in Figure 4 is obtained.

The hybrid plasmids shown in Figure 3 and Figure 4 are transformed into E. coli K 12, and the bacteria are cultured in 2 YT medium (Miller, Experiments in Molecular Genetics; 1972) until an optical density of 1 at 578 nm in the shake culture is reached. 0.1 mM IPTG (isopropyl-β-thiogalactopyranoside) is added to one portion of the bacterial culture, and thus induces the synthesis of γ-interferon for 2 hours.

The bacteria are then removed by centrifugation and disrupted by treatment with lysozyme, EDTA and ultrasound (Maniatis et al., Molecular Cloning, Cold Spring Harbor, 1982). The interferon titers of the bacterial lysates are found with a commercially available radioimmunoassay (Celltech) to be as follows:

Plasmid shown in Figure 3: $1 \times 10^5$ units per ml

Plasmid shown in Figure 4: $1.5 \times 10^5$ units per ml.

Thus, in comparison with the tac region, which is known to be excellent, the γ-interferon yield obtained with the control region according to the invention is 50% higher.

DNA sequence I (coding strand)

5' GGATCCTAAATAAATTCTTGACATTT1TTAA2TAATTTGGTATAATGT3T

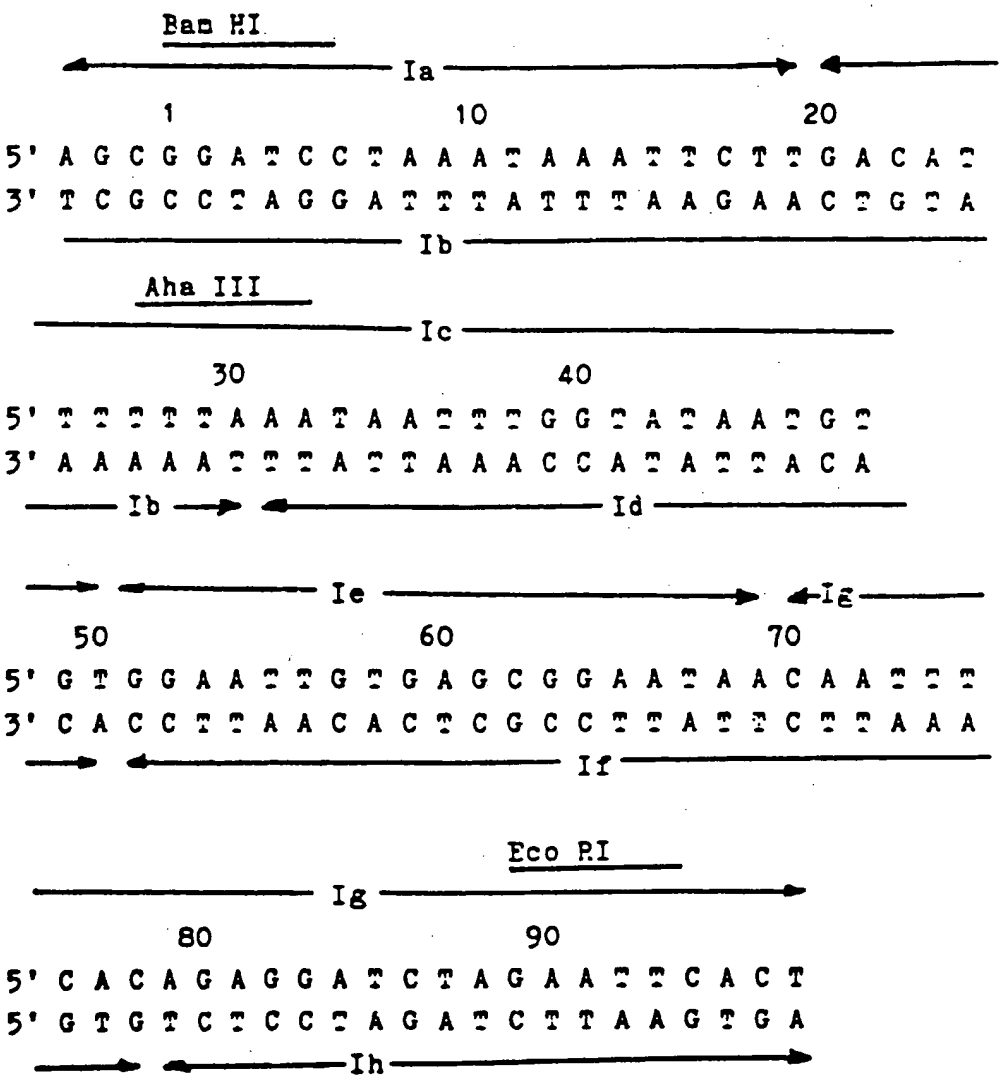
4GAATTG5GAGCG6T7ACAATT8C9C10G11G12T13TA14TT15 (ATG) 3'

1 = T or G
2 = A or C
3 = G or C
4 = G or A
5 = T or C
6 = C, GA or GAA
7 = A or C
8 = T or direct bond
9 = A or TAGA
10 = A, TTTAAA, AAGCTT or AAGCTA
11 = AG or GA
12 = A or G
13 = C or T
14 = GAA or AGC
15 = C or direct bond

DNA sequences IIa-h:

| | 1 | 2 | 5 | 6 | 7 | 8 | 10 | 11 | 12 | 13 | 14 | 15 | 3 = G |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| IIa | T | A | T | GAA | A | T | A | AG | A | C | GAA | C | 4 = G |
| b | T | A | T | GAA | A | T | TTTAAA | AG | A | C | GAA | C | 9 = A |
| c | G | C | T | GAA | A | T | TTTAAA | AG | A | C | GAA | C | |
| d | G | C | T | GAA | A | - | AAGCTT | AG | A | C | GAA | C | |
| e | G | C | C | GAA | A | - | AAGCTT | AG | A | C | GAA | C | |
| f | G | C | T | C | C | - | AAGCTT | AG | A | C | GAA | C | |
| g | G | C | T | GA | C | - | AAGCTT | AG | A | C | GAA | C | |
| h | G | C | T | GA | C | - | AAGCTA | GA | G | T | AGC | - | |

DNA sequence IIa



THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. A synthetic regulation region for the expression of heterologous genes in E. coli, which contains a promoter, a modified lac operator and a ribosomal binding site, having one or more of the following features:

a) a spacer group of 15 to 18 base-pairs is located between the -35 and the -10 regions, and

b) a spacer group of 6 to 14 base-pairs is located between the ribosomal binding site and the ATG start codon

and optionally:

c) the -35 region in the promoter has the following nucleotide sequence (coding strand)

TTGACAT or CTTGACAT,

d) the -10 region in the promoter has the following nucleotide sequence (coding strand)

GTATAAT.

2. A regulation region as claimed in claim 1 wherein the spacer group between the -35 and the -10 regions has 17 base-pairs and/or is rich in A and T.

3. A regulation region as claimed in claim 1 or claim 2, wherein the spacer group between the ribosomal binding site and the start codon has 10 base-pairs and/or is rich in A and T.

4. A regulation region as claimed in one or more of the preceding claims, wherein the ribosomal binding site is rich in purines.

5. A regulation region as claimed in one or more of the preceding claims, wherein the lac operator has the DNA sequence I or IIa as herein defined.

6. A gene structure containing a regulation region as claimed in any one of claims 1 to 5.

7. A hybrid vector containing a gene structure as claimed in claim 6.

8. E. coli containing a hybrid vector as claimed in claim 7.

9. A polypeptide expressed by E. coli as claimed in claim 8.

DATED this 16th day of October, 1989.

HOECHST AKTIENGESELLSCHAFT

WATERMARK PATENT & TRADEMARK ATTORNEYS
50 Queen Street,
MELBOURNE. VIC. 3000.
AUSTRALIA.

DBM:JMW:JZ (9.22)

FIG.1

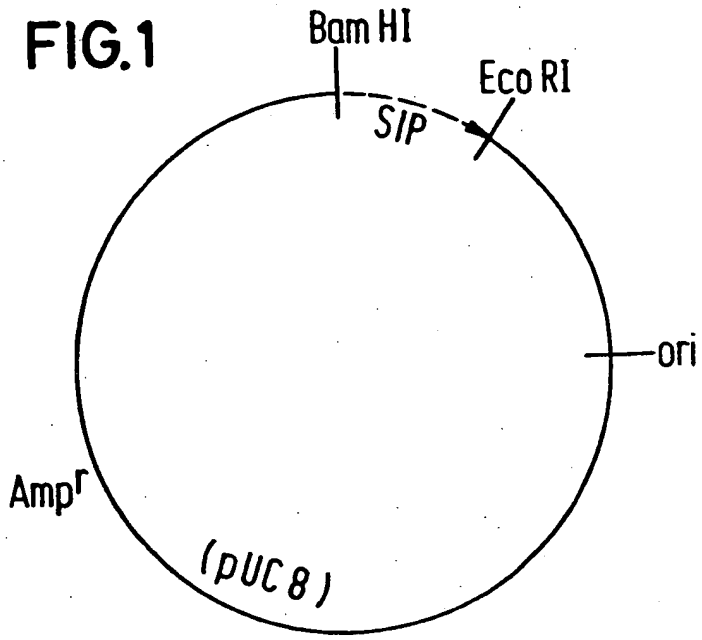

Bam HI
Eco RI
SIP
ori
Amp$^r$
(pUC 8)

FIG.2

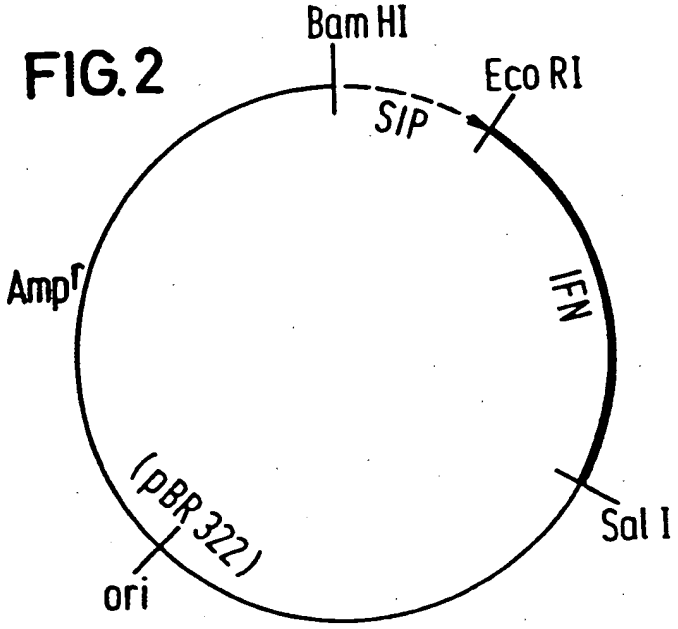

Bam HI
Eco RI
SIP
IFN
Amp$^r$
(pBR 322)
ori
Sal I

FIG.3

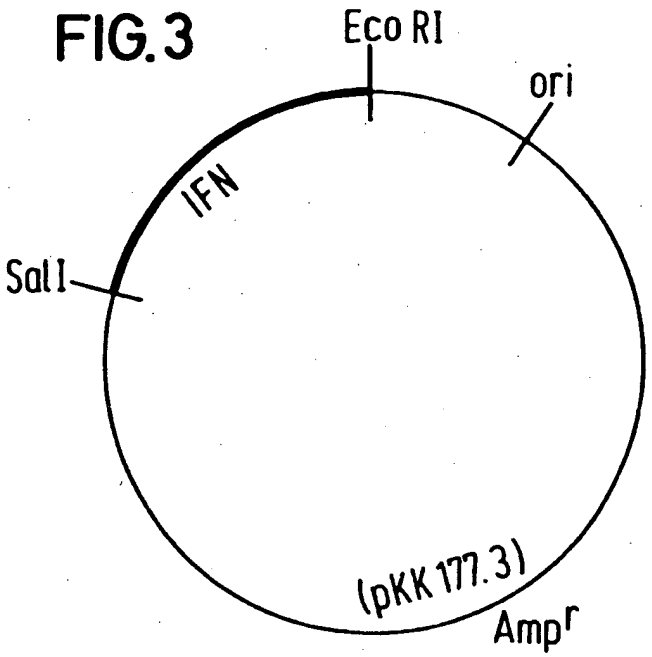
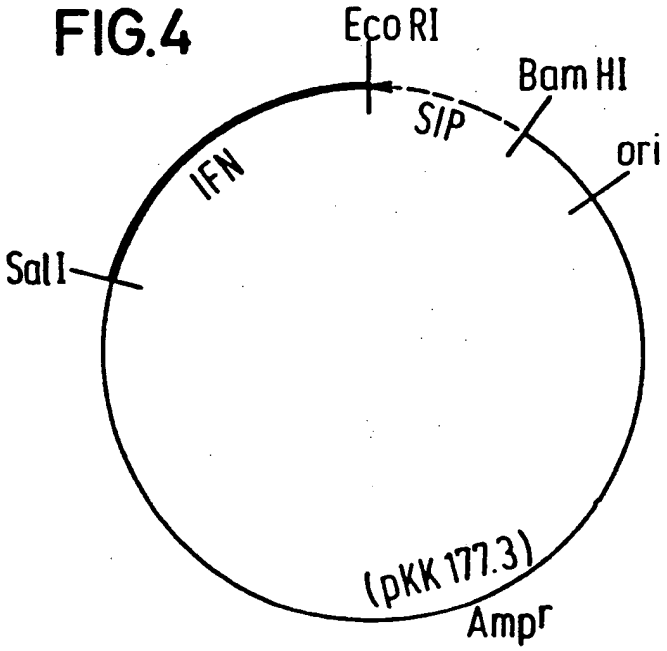

Eco RI
ori
IFN
Sal I
(pKK 177.3)
Amp^r

FIG.4


Eco RI
Bam HI
SIP
ori
IFN
Sal I
(pKK 177.3)
Amp^r